Bildretrieval mit dynamisch extrahierten Merkmalen

Odej Kao

Institut für Informatik, Universität Paderborn,
Fürstenallee 11, 33102 Paderborn
okao@uni-paderborn.de

Der aktuelle Ansatz für Bildretrieval basiert auf der Extraktion und dem Vergleich von a-priori definierten Merkmalen. Diese beschreiben das gesamte Bild und berücksichtigen einzelne Details proportional zu deren räumlicher Ausdehnung. Dies hat zur Folge, dass der Bildhintergrund den Vergleich dominiert, d.h. Bilder mit ähnlichen Objekten aber in unterschiedlichen Umgebungen werden nicht als ähnlich klassifiziert. Untersuchungen von häufig vorkommenden Anfragen in existierenden Datenbanken zeigen jedoch, dass die Mehrheit der Benutzer nach Bildern mit bestimmten Objekten, Personen, bekannten Symbolen usw. suchen möchte. Daher wird in diesem Beitrag eine Möglichkeit zur Lösung dieses Problems vorgestellt, die auf einer gezielten Definition und Beschreibung von benutzerdefinierten Regionen basiert. Diese werden dann zur Laufzeit mit allen Bildausschnitten verglichen, so dass die Ähnlichkeit zweier Bilder über die Ähnlichkeit von darin enthaltenen Regionen festgelegt. Für die Transformation der Regionen werden wavelet- und gabor-basierte Verfahren eingesetzt.

1 Einführung

Die Verwendung von a-priori extrahierten Merkmalen für Bildretrieval führt dazu, dass der Vergleich von den Eigenschaften der globalen Szene dominiert wird und Bilddetails wie Personen und Objekte oft unzureichend repräsentiert werden. Beim Suchvorgang bleiben solche Bilddetails somit unberücksichtigt, wie dies am folgenden einfachen Beispiel verdeutlicht wird.

Es sei das in Abbildung 1 dargestellte Anfragebild mit einer Person am Strand vorgegeben. Eine farbbasierte Beschreibung dieses Bildes wird durch die gelben und blauen Töne des Hintergrunds beherrscht, so dass die Merkmale der abgebildeten Person, wie etwa die Farben der Jacke oder des Gesichts, kaum ins Gewicht fallen. Analog dazu werden beispielsweise die Waveletkoeffizienten über das gesamte Bild verteilt und gewichten die Bildbereiche entsprechend ihrer Ausdehnung. Da die Person nur einen kleinen Teil einnimmt, kann sie in beiden Fällen den global durchgeführten Vergleich ganzer Bilder nicht beeinflussen. Das Ergebnis ist ein Ranking, bei dem ähnliche Strandbilder an den ersten Positionen zu finden sind. Bilder mit der Person in anderen Umgebungen, wie etwa in einem Wald, werden nicht als ähnlich identifiziert und ans Ende des Rankings einsortiert.

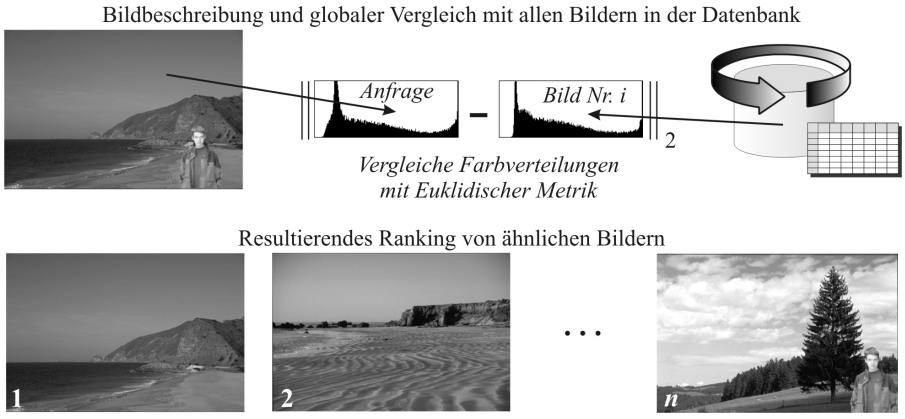

Abbildung 1: Globaler Bildvergleich mit a-priori extrahierten Merkmalen

Mit diesem standardmäßig verwendeten Ansatz ist es also nicht möglich, eine präzise Suche nach spezifischen Bilddetails durchzuführen. Eine solche Funktionalität wird von Benutzern jedoch oft erwünscht, da es der Vorgehensweise eines menschlichen Beobachters bei der Bildanalyse entspricht. Nach einer ersten Erfassung der globalen Szene findet eine Fokussierung auf interessante Bildbereiche statt, so wird beispielsweise registriert, welche Personen oder Objekte zu sehen sind. Diese Annahme wird auch durch Untersuchungen der von Journalisten eingegebenen Medienbeschreibungen gestützt [1]. Zur Realisierung dieser Vorgehensweise wird im Folgenden ein auf dynamischer Merkmalsextraktion basierender Ansatz vorgestellt.

2 Dynamisches Bildretrieval

Das Retrieval mit dynamischer Merkmalsextraktion – auch dynamisches Bildretrieval genannt – unterscheidet sich vom Standardansatz für Bildretrieval in zwei Punkten. Zum einen wird eine interaktive Bestimmung der interessanten Bildbereiche (*Regions of Interest*) seitens des anfragenden Benutzers erlaubt. Alle anderen Teile des Bildes werden ignoriert, so dass sich alle verfügbaren Beschreibungsressourcen auf den selektierten Ausschnitt konzentrieren. Zum anderen wird der Vergleich ganzer Bilder durch einen Vergleich mit einzelnen Ausschnitten ersetzt. Ein Bild wird als Treffer gewertet, wenn es einen Ausschnitt enthält, welcher der gesuchten, benutzerdefinierten Region hinreichend ähnlich ist. Diese Vorgehensweise wird grafisch in Abbildung 2 dargestellt. Das resultierende Ranking zeichnet sich dadurch aus, dass Fotos mit der angefragten Person auf den ersten Positionen zu finden sind. Die Vernachlässigung des Hintergrunds führt zur niedrigen Bewertung der restlichen Strandaufnahmen, so dass diese aus der Ergebnisliste entfernt werden.

Abbildung 2: Bildvergleich mit dynamisch extrahierten Merkmalen

Die Suche nach benutzermarkierten Regionen ist nur eine Möglichkeit für den Einsatz von Bildretrieval mit dynamischer Merkmalsextraktion. Eine Transformation der bereits gespeicherten Medien während der Anfragebearbeitung ist auch dann notwendig, wenn Attribute an die aktuelle Problemklasse angepasst werden müssen. So kann beispielsweise die Anzahl von a-priori extrahierten Waveletkoeffizienten beim Standardansatz während der Laufzeit nicht erhöht werden. Wenn also detailreiche Bilder durchsucht werden müssen, ist eine erneute Waveletzerlegung aller gespeicherten Bilder notwendig, d.h. die Waveletkoeffizienten müssen dynamisch extrahiert und verglichen werden. Zur Reduktion der notwendigen Speicherressourcen werden ferner Merkmale mit einem hohen Speicherbedarf aber einem moderaten Berechnungsaufwand erst auf Verlangen extrahiert. Kantenbilder werden beispielsweise selten als Merkmale gebraucht, benötigen aber annähernd genauso viel Speicher wie die Bildrohdaten. Da die Erstellung eines Kantenbildes nur den Bruchteil einer Sekunde dauert, kann die Kantenextraktion bei Bedarf und während der Laufzeit durchgeführt werden.

In Anbetracht des momentanen Entwicklungsstands von Bilddatenbanken ist jedoch die Umsetzung einer Detailsuche das wichtigste Ziel. Aus diesem Grund werden im Folgenden existierende Methoden für die Beschreibung und den Vergleich von benutzermarkierten Regionen angepasst und neue Methoden für Retrieval mit dynamischer Merkmalsextraktion entwickelt.

2.1 Testreihen für die Bewertung der Retrievalqualität

Für die Leistungsmessung der Methoden für dynamisches Bildretrieval wird eine Reihe von Fotoserien verwendet, die synthetische Bilder, Bilder mit einkopierten Objekten und reale Aufnahmen beinhalten. Diese wurden so gewählt, dass eine stufenweise Komplexitätssteigerung und somit eine Bewertung der Retrievalqualität in verschiedenen Umgebungen ermöglicht wird. Im Einzelnen handelt es sich um:

- Abzeichen und Verkehrsschilder: Standardbilder (75) von Abzeichen und Logos verschiedener Organisationen sowie Verkehrsschilder wurden auf eine einheitliche Größe skaliert. Sie weisen eine Fülle von Details auf, so dass mit dieser Serie die generellen Erkennungsfähigkeiten der neuen Retrievalmethoden untersucht werden sollen.
- Landschaftsaufnahmen: Diese Fotoserie besteht aus 114 Aufnahmen von Nationalparks, so dass viele verschiedene Szenarien und Farben vorhanden sind. In 14 Bildern wurden nachträglich reale Objekte – in diesem Fall eine Person – einkopiert, die anschließend als Anfrageregion dient. Die Person wurde dabei auf 50% und 80% verkleinert, um 14° und 90° rotiert, horizontal und vertikal gespiegelt und teilweise verdeckt in die Bilder eingefügt, so dass realistische Objektanordnungen simuliert werden können.
- Waldserie: Diese realen Aufnahmen zeigen zwei Personen in Waldszenarien und zeichnen sich durch verschiedene Lichtverhältnisse, Abstände der Personen zur Kamera und ähnliche Hintergrundszenarien aus. Die abweichende Körperhaltung und Gestik erschweren die Objektsuche und -erkennung zusätzlich, so dass mit diesen Aufnahmen Bildretrieval hoher Komplexität simuliert wird.

Die Anzahl der betrachteten Bilder erscheint auf den ersten Blick gering. Allerdings muss beachtet werden, dass bei der verwendeten Schrittgröße von fünf Pixel etwa 3923 Bereiche pro modifiziertem Bild und 2099 Bereiche pro Waldbild zu analysieren sind. Dies entspricht dem Bestand einer Standardbilddatenbank mit annähernd 531257 Bildern.

2.2 Mustervergleich (*Template Matching*)

Eine der grundlegenden Techniken für die Erkennung von Objekten in Bildern und Videosequenzen basiert auf einem direkten Vergleich der Anfrageregion mit allen möglichen Sektionen im Zielbild bzw. -video. Dabei besteht eine große Vielfalt von Vergleichsoperationen, Merkmalen sowie Auswahlmöglichkeiten. Die einfachste Form beruht auf einer unmittelbaren Verknüpfung der Pixel in der Region mit den korrespondierenden Pixel des darunter liegenden Bildbereichs, beispielsweise durch eine absolute Subtraktion der Farbwerte. Die Region wird in der Regel durch ein Rechteck minimalen Umfangs dargestellt und pixelweise über das Bild geschoben. Die Verknüpfungsoperation wird bei jeder neuen Position durchgeführt und die Ergebnisse anschließend bewertet. Daraus ergibt sich ein Ähnlichkeitsmaß für die Übereinstimmung der beiden Regionen.

Beim Einsatz des Mustervergleichs zum dynamischen Bildretrieval werden vom Benutzer die interessanten Regionen markiert und anschließend in einem Rechteck minimalen Umfangs zusammengefasst. Während der Laufzeit werden die darin enthaltenen Pixel mit einer Auswahl von Bildausschnitten – gemäß einer benutzerdefinierten Schrittweite – verglichen. Die Verknüpfungsoperation kann die

Farbe einzelner Pixel, die Farbverteilung in der Region sowie die Position der Kanten berücksichtigen. Daraus ergeben sich folgende Verfahren:

- Farbbasierter Vergleich korrespondierender Pixel,
- Regionenvergleich mit Histogrammen und statistischen Farbmomenten,
- Pixelvergleich mit histogrammbasierter Vorverarbeitung und
- Kantenbasiertes Bildretrieval.

Beim farbbasierten Vergleich werden die Farbwerte der Anfrageregion von den korrespondierenden Pixel des aktuellen Bildausschnitts subtrahiert und die Beträge zu einer Summe addiert. Dieser Vorgang wird auch für die anderen Bildsektionen wiederholt, so dass der Ausschnitt mit der kleinsten Summe den ähnlichsten Bereich markiert. Dieser Wert wird mit dem gesamten Bild assoziiert und dient als Grundlage für die Erstellung des Ähnlichkeitsrankings. Leistungsmessungen mit den Beispielfotoserien ergaben eine Recallquote von 100%, wenn nach Verkehrsschildern oder Logos gesucht wurde. Beim Retrieval der modifizierten Bilder diente das eingefügte Objekt auch als Anfrageregion und wurde mit einer Schrittweite von fünf Pixel über das Bild geschoben. Der Verzicht auf die Wiederholung des Retrievals mit verschiedenen Skalierungen und Rotationen der Anfrageregion führte zu einer niedrigen Erkennungsquote von 57.14%, d.h. es wurden acht der 14 möglichen Bereiche korrekt markiert und unter den ersten 32 Treffern zurückgegeben. Eine weitere Verschlechterung der Ergebnisse erfolgte beim Betrachten der Serie mit den realen Bildern, da hier nur drei der möglichen 23 Objekte (13.04%) gefunden wurden. Die hier durchgeführte Korrelation kann auf unterschiedliche Arten, beispielsweise durch eine Faltung im Frequenzraum, realisiert werden. Auf vergleichende Leistungsmessungen wurde allerdings verzichtet, da an dieser Stelle ausschließlich die prinzipielle Eignung der entwickelten Verfahren für das dynamische Bildretrieval untersucht wird. Aus dem gleichen Grund werden auch keine benutzerdefinierten Anpassungen, wie etwa die Einführung von Gewichten, erlaubt, da dadurch die Abhängigkeit der Recallquoten von den aktuellen Testbildern stark zunimmt.

Bei histogrammbasierten Verfahren wird die Information über die Position der Farben in den untersuchten Regionen teilweise – wie bei Blob-Histogrammen – oder gänzlich vernachlässigt, so dass ausschließlich die Farbenhäufigkeit in die Ähnlichkeitsberechung eingeht. Nach einer starren Quantisierung des Farbraums werden Histogramme mit jeweils 256 Einträgen bestimmt und die einzelnen Komponenten voneinander subtrahiert. Die aufsummierten absoluten Differenzen dienen als Ähnlichkeitsmaß. Der Bildausschnitt mit der niedrigsten Summe wird als ähnlichster Bereich markiert und repräsentiert das gesamte Bild bei der Erstellung des Rankings. Eine weitere Möglichkeit zur farbbasierten Repräsentation der Anfrageregion und der Bildausschnitte basiert auf der Verwendung von statistischen Farbmomenten. Die Farbverteilung wird durch die Durchschnittsintensität, Standardabweichung und Asymmetrie der Bildkanäle approximiert, so dass ein Merkmalsvektor mit neun Einträgen als Vergleichsgrundlage dient. Diese kompakte Darstellung beschleunigt den Retrievalprozess.

Die Anwendung der histogrammbasierten Verfahren auf die Testbilder zeigte, dass sich die Vernachlässigung der exakten Farbposition nachteilig auf die Retrievalqualität auswirkt. Durch einen unmittelbaren Vergleich der korrespondierenden Histogramme wurden lediglich 14.29% der Objekte in den modifizierten Abbildungen und 26.09% der relevanten Waldfotos gefunden. Die Approximation der Farbverteilung mit statistischen Farbmomenten ermöglichte keinen nennenswerten Fortschritt, da nur 21% der Bilder in den beiden Fotoserien korrekt identifiziert wurden.

2.3 Retrievalgüte bei direktem Pixelvergleich

Eine Reihe von Problemen kann mit einem Mustervergleich zufriedenstellend gelöst werden, wenn die Voraussetzung einer begrenzten und wohl-definierten Menge von Zielmustern – wie zum Beispiel zulässige Verkehrszeichen samt möglicher Variationen – erfüllt ist. Da dies für Bilddatenbanken im Allgemeinen nicht zutrifft, kann kein anwendungsspezifisches Wissen in den Retrievalprozess eingebunden werden. Aus diesem Grund wurden bei der Suche nach der Kopie einer Anfrageregion zwar hohe Recallquoten erzielt, nicht jedoch bei der Suche nach ähnlichen Objekten in einem realen Bestand. Für diese Art des Bildretrievals ist die Fixierung auf Farben und deren Position nicht ausreichend. Ferner verschlechtern sich die Ergebnisse, wenn zusätzliche Störungen im verwalteten Bildbestand auftreten, zum Beispiel:

- Aufnahmedefekte: Unterschiedliche Aufnahmebedingungen führen zu Variationen in den Farbwerten, dem Bildkontrast und der Bildschärfe.
- Skalierungen: Die Abbildungsgröße eines Objekts hängt von Faktoren wie Entfernung von der Kamera ab, so dass beim Vergleich verschiedene Skalierungen berücksichtigt werden müssen.
- Rotationen: Analog zu den unterschiedlichen Skalierungen kann das Objekt einen beliebigen Rotationswinkel annehmen. Die in verschiedenen Größen erstellten Vorlagen müssen für alle zulässigen Winkel kopiert, rotiert und als Schablone verwendet werden.
- Positionierung der Schablone: Die eingesetzte Schrittweite stellt einen Kompromiss zwischen Rechengeschwindigkeit und Retrievalgenauigkeit dar. Idealerweise wird jede mögliche Position untersucht, der hierzu notwendige Rechenaufwand ist jedoch in der Regel nicht vertretbar. Größere Schrittweiten ermöglichen zwar kürzere Antwortzeiten, die Verschiebung der untersuchten Regionen verfälscht jedoch die Ähnlichkeitswerte.

Die Komplexität eines direkten Pixelvergleichs wird erhöht, wenn neben den Variationen in der Skalierung und Rotation noch weitere Abweichungen wie etwa in der Objektform oder -textur zugelassen sind. Unterschiedliche Kleidung der abgebildeten Person führt zu wesentlich höheren Differenzen und beeinträchtigt das Ähnlichkeitsranking. Der gleiche Effekt ergibt sich, wenn die Farben und Texturen aufgrund der Aufnahmebedingungen verändert wurden. Ferner werden zu viele, unter

Umständen bedeutungslose Pixel untersucht, deren Differenzen jedoch vollständig im Ergebnis integriert werden.

Diese gleichmäßige Berücksichtigung kann durch eine partielle Betrachtung ersetzt werden: Transformationsbasierte Methoden wandeln die Bereiche in einen Merkmalsraum um, in dem signifikante Bildbereiche hervorgehoben und vorrangig beschrieben werden. Zwei Methoden dieser Art, die auf der Bildbeschreibung mit Waveletkoeffizienten sowie mit Gabor-Wavelet-Merkmalen basieren, werden im Folgenden entwickelt.

3 Bildretrieval mit dynamischer Extraktion waveletbasierter Merkmale

Bei einem waveletbasierten Bildretrieval werden die Bilder mit der Wavelettransformation zerlegt und durch eine Reihe von Koeffizienten beschrieben. Deren Anzahl wird empirisch bestimmt und liegt üblicherweise zwischen 32 und 64 Koeffizienten [2]. Die Koeffizienten sind über das gesamte Bild verteilt, so dass einzelne Objekte kaum erfasst und somit unzureichend repräsentiert werden. Bei den hier entwickelten Verfahren werden deswegen nur die vom Benutzer markierten Regionen transformiert und die Koeffizienten so eingesetzt, dass die interessanten Bereiche möglichst präzise approximiert werden.

Die Anfrageregion und die Ausschnitte in den Zielbildern werden mit der schnellen Wavelettransformation (FWT) zerlegt [3]. Für das dynamische Retrieval werden die Approximationskoeffizienten verwendet, da diese die charakteristischen Informationen eines Bildes repräsentieren und deren Eignung für die Beschreibung und Erkennung wichtiger Regionen experimentell bestätigt wurde. Analoge Aussagen wurden ferner von STARCK ET AL. [4] und von OREN ET AL. [5] getroffen. Der Vergleich des Anfragebereichs mit Bildausschnitten erfolgt nach folgendem Algorithmus:

```
for alle Bilder begin
   for alle Zeilen gemäß Schrittweite begin
      for alle Spalten gemäß Schrittweite begin
         Bestimme und zerlege den aktuellen Bildausschnitt mit der FWT
         Bewerte die Ähnlichkeit zur Anfrageregion auf Basis der
         Approximationskoeffizienten und des gewählten Trefferkriteriums
      end
   end
   Speichere den Wert des ähnlichsten Ausschnitts für das aktuelle Bild
end
Sortiere die Bilder nach aufsteigenden Ähnlichkeitswerten
```

Im weiteren Verlauf werden vier waveletbasierte Verfahren für dynamisches Bildretrieval eingeführt und somit vier verschiedene Trefferkriterien vorgeschlagen.

Die CC-Suche (*Corresponding Coefficients*) beruht auf der Annahme, dass sich bei ähnlichen Regionen auf demselben Zerlegungslevel und an derselben Position auch ähnliche Waveletkoeffizienten befinden. Während der Suche wird die Ähnlichkeit korrespondierender Koeffizienten untersucht.

Gesamtähnlichkeitswert = 0
Runde die Koeffizienten für die Anfrageregion auf
for alle Zerlegungsebenen begin
Runde die Ausschnittskoeffizienten auf
Berechne die Differenz zwischen Anfrage- und Ausschnittskoeffizienten
Zähle die Nullen in der Differenz
Addiere diesen Wert zum Gesamtähnlichkeitswert
end

Mit einer zusätzlichen Schwelle werden die nahe bei Null liegenden Werte auf Null abgebildet und somit die Anzahl der zu berücksichtigenden Punkte erhöht.

In dem Algorithmus DALC-Suche (*Difference of Absolute Largest Coefficients*) werden die Differenzen einer Auswahl von betragsgrößten Koeffizienten betrachtet, welche den Abstand des untersuchten Signals zu den skalierten und translatierten Basiswavelets angeben. Diese Abhängigkeit der Bildinformation von den Beträgen der Waveletkoeffizienten wird bereits in mehreren Verfahren zur Bestimmung der Ähnlichkeit ganzer Bilder ausgenutzt [6,7].

Nach der Waveletzerlegung der Anfrageregion werden die betragsgrößten Waveletkoeffizienten und ihre Positionen identifiziert. Während der Anfragebearbeitung wird der aktuelle Bildausschnitt im Zielbild ebenfalls mit der Wavelettransformation zerlegt. Anschließend werden aus dem Ergebnis die Koeffizienten an den durch die Analyse der Anfrageregion ermittelten Positionen extrahiert. Die Subtraktion und Betragsbildung zwischen den korrespondierenden Koeffizienten des Ausschnitts und der Anfrageregion ergeben schließlich den Ähnlichkeitswert für die aktuelle Zerlegungsebene. Dieses Vorgehen wird für die Ebenen mit feinerer Auflösung wiederholt und die Beträge der einzelnen Ähnlichkeitswerte zu einem Gesamtwert aufaddiert. Wichtige Parameter sind die Anzahl der betrachteten betragsgrößten Koeffizienten sowie die Gewichtung der Dekompositionsebenen. Letzteres führt dazu, dass die einzelnen Zerlegungsebenen im Sinne der Multiskalenanalyse unterschiedliche Detaillierungsgrade eines Bildes repräsentieren.

Bei der DALC-Suche werden 256 Graustufen betrachtet, so dass eine Bildtransformation in der Vorverarbeitung erfolgen muss. Bei der RGB-Suche wird hingegen die gesamte Farbinformation verwendet, indem die Wavelettransformation auf jeden Farbkanal angewendet wird. Analog zur DALC-Suche werden damit die betragsgrößten Koeffizienten pro Kanal bestimmt und verglichen.

Der Á trous-Suche liegt die so genannte Wavelettransformation mit Löchern (fr. *la décomposition Á trous*) – im Folgenden als Á trous-Transformation bezeichnet – zugrunde. Bei der Á trous-Transformation wird keine Unterabtastung durchgeführt, so dass die Anzahl der Detail- und Approximationskoeffizienten pro Zerlegungsebene mit der Pixelanzahl im Originalbild identisch ist, man spricht von einer redundanten

Transformation. Die Waveletzerlegung wird mit dem gesamten Zielbild durchgeführt und die transformierten Bilder werden als a-priori extrahierte Merkmale in der Datenbank abgelegt. Damit entfällt die Transformation der zu untersuchenden Bildbereiche während der Laufzeit. Lediglich die Zerlegung der Anfrageregion und die Vergleiche der korrespondierenden Koeffizienten müssen während der Anfragebearbeitung ausgeführt werden. Diese Zeitersparnis eröffnet Möglichkeiten für den Einsatz des dynamischen waveletbasierten Retrievals in traditionellen Datenbanksystemen. Beim verwendeten Algorithmus für die Á trous-Transformation [8] wird keine Unterteilung in horizontale, vertikale und diagonale Details vorgenommen. Stattdessen wird zwischen den Approximationen, die für das Retrieval verwendet werden, und den Details eines Bildes unterschieden. Letztere werden aus der Differenz der Approximationen auf zwei aufeinander folgenden Ebenen erhalten. Der Abstand zwischen den herangezogenen Punkten aus der Vorebene vergrößert sich jeweils um den Faktor Zwei. Dazwischen befinden sich die namengebenden Löcher. Dieser Prozess wird mit jeder neuen Ebene fortgesetzt. Eine Reihe von Experimenten zeigte, dass der bislang verwendete direkte Vergleich von betragsgrößten Koeffizienten zur Bestimmung der Ähnlichkeit zwischen der Anfrageregion und dem aktuellen Bildausschnitt im Fall der Á trous-Zerlegung zu unbefriedigenden Ergebnissen führt. Aus diesem Grund wird ein alternatives Vergleichskriterium gewählt, das einige charakteristische Eigenschaften, wie minimale und maximale Koeffizienten, Standardabweichung und Mittelwert aller Koeffizienten, in einem Ähnlichkeitswert zusammenfasst.

3.1 Bewertung der waveletbasierten Methoden

Zur Validierung der grundlegenden Erkennungsfähigkeit der vorgestellten Verfahren für dynamisches Bildretrieval wurden zunächst Experimente mit den 75 synthetischen Bildern – Verkehrsschildern und Abzeichen – durchgeführt. Die durchgeführten Messungen zeigten, dass das Anfragebild unabhängig vom gewählten Retrievalverfahren stets als erster Treffer zurückgegeben wurde. Die weiteren Treffer konnten ebenfalls als ähnlich klassifiziert werden, da die äußere Form übereinstimmte. Insbesondere wurden die drei Varianten eines Objekts an den ersten drei Positionen gefunden.

Für die Messungen der Retrievalqualität im Fall der Fotoserien wurde die einkopierte Person als Anfrageregion verwendet. Mit dieser Vorlage wurden Suchvorgänge entsprechend der Vorgehensweise des dynamischen Bildretrievals gestartet. Dabei wurden die Ausschnitte im Zielbild untersucht, welche in jeder fünften Position einer Zeile und jeder fünften Spalte beginnen. Die Ähnlichkeit eines Bildes mit der Anfrageregion wurde durch den Wert desjenigen Ausschnitts dargestellt, bei dem die größte Übereinstimmung mit der Anfrageregion festgestellt wurde. Dieser Bereich wurde durch ein Rechteck markiert, so dass zwischen tatsächlichen und zufälligen Treffern differenziert werden konnte. Diejenigen Ergebnisbilder, welche die gesuchte Person zwar enthalten, diese aber nicht mit dem Rechteck markiert war, wurden als nicht erkannt gewertet. Die maximale

Zerlegungsebene hängt von der Größe der Anfragevorlage ab. Bei der aktuellen Testreihe wurden drei Ebenen erzeugt und mit 48 bzw. 448 Koeffizienten beschrieben. Diese Größen wurden willkürlich gewählt und stellvertretend für eine Bildanalyse mit einer kleinen und einer großen Anzahl von Koeffizienten eingesetzt. Auf eine Gewichtung wurde verzichtet, da diese an die aktuelle Bildklasse angepasst werden muss und die Vergleichbarkeit der Ergebnisse beeinträchtigt.

Die erste Phase dieser Testreihe beinhaltete die Suche nach 14 modifizierten Bildern, die sich in einer Menge von 114 Bildern befanden. Die DALC- und die RGB-Suche erzielten mit jeweils zwölf erkannten Objekten – entspricht einer Recallquote von 85.7% – die besten Ergebnisse. Danach folgte die Á trous-Suche mit 78.6% und anschließend die CC-Suche mit 71.4%. Bei keiner der vier Suchfunktionen wurden alle 14 Objekte erkannt. Ferner lieferte keine der Funktionen alle gefundenen Objekte innerhalb der ersten 32 Treffer zurück. Mit 71.4% erreichte die RGB-Suche hierbei das beste Ergebnis, gefolgt von der DALC-, der Á trous- und schließlich der CC-Suche. Eine Darstellung der zurückgegebenen Bilder am Beispiel der RGB-Suche findet sich in Abbildung 3. Aus Übersichtlichkeitsgründen sind nur die ersten 16 Treffer angezeigt.

Abbildung 3: Die 16 besten Treffer bei der RGB-Suche, von links nach rechts und von oben nach unten

In der zweiten Phase wurden die waveletbasierten Verfahren auf die realen Bilder der Waldserie angewendet, um die Retrievalqualität unter allgemeinen Bedingungen zu untersuchen. Dabei muss berücksichtigt werden, dass die Lichtverhältnisse,

Abstände zur Kamera sowie die Körperhaltung und Gestik der abgebildeten Personen von Bild zu Bild variieren und somit die Erkennung erschweren. Dementsprechend fiel die Recallquote ab, so dass – im Fall der RGB-Suche – maximal 43.5% aller gewünschten Bilder tatsächlich auch ermittelt wurden. Die erzielten Ergebnisse mit den anderen Verfahren lagen unter dieser Recallquote, so wurden mit der DALC- und der Á trous-Suche lediglich 21.7% der relevanten Bilder erkannt. Eine ausführliche Darstellung dieser Ergebnisse und der waveletbasierten Verfahren für dynamisches Bildretrieval findet sich in [9].

4 Dynamisches Bildretrieval mit Gabor-Wavelets

Merkmale, die auf einer Bildtransformation mit Gaborfunktionen als Mutterwavelets (*Gabor-Wavelets*) basieren, werden aufgrund ihrer Invarianzeigenschaften gegenüber affinen Transformationen und Beleuchtungsänderungen häufig zur Objektrepräsentation eingesetzt. Insbesondere in Bereichen wie Klassifikation von Objekten [10], Gesichtserkennung oder Gesichtsverfolgung in Videosequenzen [11] werden solche Merkmale berücksichtigt. Diese erfolgreiche Verwendung der Gabor-Wavelets dient als Motivation, die Eignung dieser Merkmale für dynamisches Bildretrieval zu untersuchen. Dazu werden im aktuellen Abschnitt zwei Verfahren für dynamisches Bildretrieval basierend auf zweidimensionalen Gabor-Wavelets entwickelt. Im ersten Fall wird ein unmittelbarer Vergleich der Filterantworten von markanten, manuell ausgewählten Kernbereichen durchgeführt. Bei dem zweiten Ansatz werden so genannte Gabor-Wavelet-Netzwerke [11] für das dynamische Bildretrieval angepasst und eingesetzt. Diese optimieren die Suche, indem sie Abweichungen hinsichtlich der Skalierung, Orientierung und Pixelkorrespondenz ausgleichen.

Die entscheidenden Impulse für die Kombination einer Wavelettransformation mit Gaborfunktionen als Mutterwavelets resultieren aus den Ergebnissen biologischer Untersuchungen des primären visuellen Kortex [12]. Seit dieser Zeit wurde eine intensive – experimentelle und theoretische – Forschung betrieben, welche in ein umfangreiches Wissen über die so genannten einfachen Zellen (*Simple cells*) resultiert. Das Antwortverhalten dieser Zellen auf visuelle Reize kann folgendermaßen modelliert werden:

$$\psi_{k,\theta}(x,y) = \frac{k^2}{\sigma^2}\exp\left[-\frac{k^2k^2}{2\sigma^2}\left((x\cos\theta + y\sin\theta)^2 + (-x\sin\theta + y\cos\theta)^2\right)\right] \times \exp\left[-ik^2(x\cos\theta + y\sin\theta)\right] - \exp\left[-\frac{\sigma^2}{2}\right] \quad (1)$$

Dabei gibt der Parameter Θ die Orientierung an. Die Wavelettransformation eines Bildes b mit einer Gaborfunktion kann somit dargestellt werden als Faltung:

$$J_{k,\theta}(x,y) = \iint b(x',y')\,\psi_{k,\theta}(x-x',y-y')\,dx'\,dy' \qquad (2)$$

Eine detaillierte Beschreibung und Analyse dieser Zusammenhänge findet sich unter anderem in [13,14]. Die Parameter der Gabor-Wavelets werden dabei an die neuropsychologischen Daten der einfachen Zellen sowie an die Wavelettheorie angepasst, so dass eine stabile Bildrekonstruktion durch eine lineare Superposition der Gabor-Wavelets ermöglicht wird.

4.1 Dynamisches Bildretrieval durch Vergleich korrespondierender Jets

Die Ergebnisse der Faltung von signifikanten Bildbereichen mit den vorgestellten Gabor-Wavelets werden als Beschreibungsmerkmale für die Bildrepräsentation und den Vergleich eingesetzt. Für diesen Zweck werden wichtige Kernbereiche vom Benutzer manuell ausgewählt und durch die Positionierung eines quadratischen Fensters markiert. Entsprechend Gleichung (2) werden die Faltungsergebnisse für den ausgewählten Bereich bezüglich einer vorgegebenen Welligkeit und einer entsprechenden Orientierung berechnet und als Merkmale gespeichert. Dazu wird der Parameterraum durch n_k Welligkeiten und n_d Orientierungen diskretisiert. Die möglichen Welligkeiten k_i und Orientierungen Θ_j werden kombiniert und zur Berechnung der Ergebnisse $J_{ki,\Theta j}$ eingesetzt. Dementsprechend ergeben sich für jeden markierten Bereich $n_k n_d$ Werte, die in einem Vektor – genannt Jet – zusammengefasst werden

$$\begin{aligned} q_i &= \left(J_{k_0,\theta_0},\cdots,J_{k_0,\theta_{n_d-1}},J_{k_1,\theta_0},\cdots,J_{k_1,\theta_{n_d-1}},\cdots,J_{k_{n_k-1},\theta_0},\cdots,J_{k_{n_k-1},\theta_{n_d-1}}\right) \\ f_i &= \left(\tilde{J}_{k_0,\theta_0},\cdots,\tilde{J}_{k_0,\theta_{n_d-1}},\tilde{J}_{k_1,\theta_0},\cdots,\tilde{J}_{k_1,\theta_{n_d-1}},\cdots,\tilde{J}_{k_{n_k-1},\theta_0},\cdots,\tilde{J}_{k_{n_k-1},\theta_{n_d-1}}\right) \end{aligned} \qquad (3)$$

Die Beträge der komplexwertigen Antworten der Wavelettransformation mit den Gaborfunktionen als Basiswavelets werden für den Vergleich eingesetzt [14,15]. Die Ähnlichkeitsbetrachtung der ausgewählten Regionen mit allen Bereichen der Bilder in der Datenbank kann somit auf eine Distanzberechnung zwischen den modifizierten korrespondierenden Jets zurückgeführt werden. Jede benutzerdefinierte Markierung wird durch die Ausdehnung des Kernbereichs, die Anzahl der zu untersuchenden Welligkeiten n_k und Orientierungen n_d sowie den Winkel der ersten Orientierung festgelegt. Anschließend werden für alle m Bereiche die Jets q_0, q_1, …, q_m bestimmt und dienen fortan als Vergleichsgrundlage. Während der Laufzeit wird eine Anzahl von Abschnitten der gespeicherten Bilder – gemäß der gewählten Schrittweite – in der gleichen Art und Weise bearbeitet. In jedem Ausschnitt werden die Jets für die identischen Kernbereiche wie in der Anfrageregion bestimmt. Für den aktuellen Bildausschnitt a_k ergeben sich demnach die Vektoren f_0, f_1, …, f_m. Anschließend wird der Abstand zwischen den korrespondierenden Jets der Anfrageregion q und des aktuellen Ausschnitts a_k mittels der Euklidischen Distanz berechnet. Zur Hervorhebung der relevanten Eigenschaften wird jedoch der vollständige

Elementvergleich durch den Vergleich der Antworten mit der ähnlichsten Orientierung ersetzt. Für ein Paar (q_i,f_i), i=1,2, …, m gilt also

$$d(q_i, f_i) = \min_{r=0}^{n_d-1} \sqrt{\sum_{j=0}^{n_k-1} \left| J_{k_j,\theta_r} - \tilde{J}_{k_j,\theta_r} \right|^2} \qquad (4)$$

Die Ähnlichkeit $S(q,a_k)$ zwischen der Anfrageregion q und dem aktuellen Ausschnitt a_k ist somit umgekehrt proportional zur Summe der Differenzen aller untersuchten Jets. Die Relevanz des gesamten Bildes für die vorliegende Anfrage wird durch den Ähnlichkeitswert derjenigen Region festgelegt, welche bei der Untersuchung die größte Übereinstimmung aufweist.

Für die Qualitätsmessungen beim dynamischen Jet-basierten Bildretrieval wurde eine Diskretisierung des Parameterraums in n_k=5 Frequenzstufen und n_d=8 Orientierungen vorgenommen. Somit werden 40 Faltungsergebnisse pro Kernbereich ermittelt und in einem Jet zusammengefasst. Das Datenfenster erstreckt sich über 20 Pixel und berücksichtigt nur eine Skalierung. Analog zu den vorherigen Messungen wird eine Schrittweite von fünf Pixel gewählt und die ähnlichsten Bereiche mit einem Rechteck gekennzeichnet. Die Anfrageregion wird mit sieben Markierungen versehen, die manuell auf signifikanten Bereichen positioniert wurden.

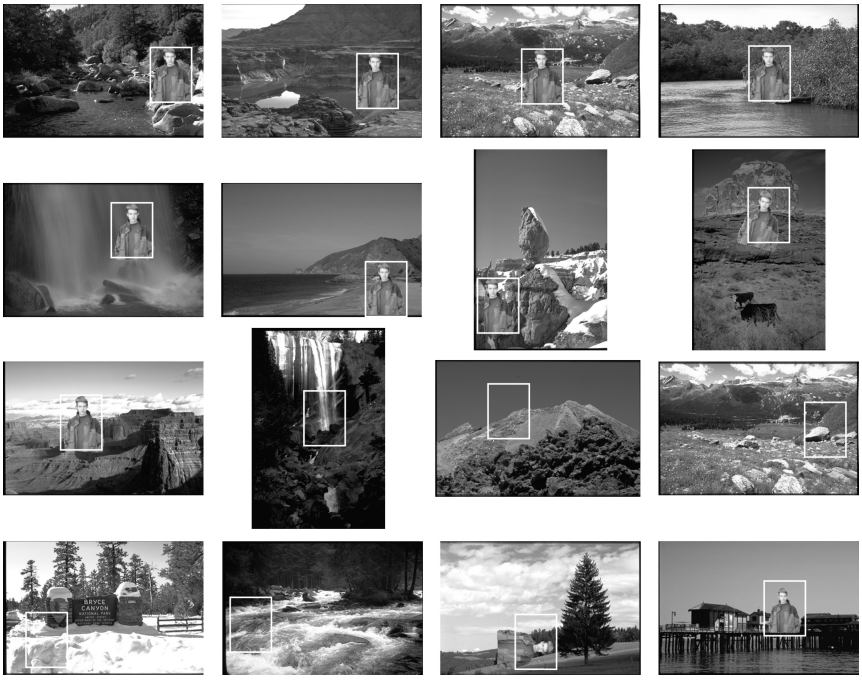

Abbildung 4: Erste 16 Treffer des Jet-basierten Retrievals in den modifizierten Landschaftsaufnahmen

Die Suche in den Verkehrschildern und Logos ergab erneut eine 100%-ige Recallquote. Bei der Serie mit modifizierten Landschaftsaufnahmen wurden 13 der 14 möglichen Bilder korrekt identifiziert und unter den ersten 32 Treffern zurückgegeben. Lediglich die vertikal gespiegelte Vorlage wurde nicht erkannt. Dieses Ergebnis entspricht einer Recallquote von 92.8% und übertrifft alle bisher vorgestellten Verfahren für dynamisches Bildretrieval. Durch die manuelle Bestimmung der Kernbereiche wird ein individuelles Kontextwissen in die Suche eingebunden, die bei den farb- und waveletbasierten Methoden nicht vorhanden war. Damit können gezielt wichtige Elemente, wie zum Beispiel das Gesicht, hervorgehoben werden. Die ersten 16 Treffer sind in Abbildung 4 zusammengefasst.

Diese Recallquote konnte im Fall der realen Aufnahmen nicht erreicht werden, da hier der unmittelbare Vergleich korrespondierender Jets durch die abweichende Körperhaltung gestört wurde. Dennoch wurden 13 der 23 relevanten Bilder korrekt identifiziert. Die entsprechende Recallquote von 56.5% lag somit höher als die mit den farb- und waveletbasierten Methoden erzielten Ergebnissen. Da sowohl die Treffer als auch die nicht erkannten Bereiche weitestgehend mit den Ergebnissen des Retrievals unter Verwendung von Gabor-Wavelet-Netzwerken übereinstimmten, wird auf eine separate Ergebnispräsentation und -analyse an dieser Stelle verzichtet.

4.2 Dynamisches Bildretrieval mit Gabor-Wavelet-Netzwerken

Gabor-Wavelet-Netzwerke (GWN) wurden von KRÜGER und SOMMER im Kontext der Gesichtsverfolgung in Videosequenzen [11] vorgeschlagen. Der Benutzer gibt die Anzahl aber nicht die Position der beschreibenden Gabor-Wavelets vor. Diese werden durch einen Optimierungsprozess so über die Anfrageregion verteilt, dass die charakteristischen Bereiche vorrangig repräsentiert werden und stärker gewichtet in die Suche eingehen. Im Vergleich zur Jet-basierenden Darstellung werden weniger Beschreibungsressourcen benötigt, die jedoch aufgrund der genaueren Positionierung eine verbesserte Bildrekonstruktion ermöglichen. Allerdings ist etwaig vorhandenes Kontextwissen über die markierte Anfrageregion nur beschränkt nutzbar, da der Benutzer – aufgrund der automatisierten Positionierung der Gabor-Wavelets – keine individuelle Einschätzung bezüglich signifikanter Bildbereiche angeben kann.

Es seien ψ_{ni} mit i=1,2 …, N eine Menge von Gabor-Wavelets und b ein mittelwertfreies Bild. Es seien ferner w_i Gewichte und n_i Parameter, die so gewählt sind, dass die Fehlerenergie E minimal wird. Dann definieren die Vektoren

$$\Psi = \left(\psi\left(s_x^1, s_y^1, \theta^1, c_x^1, c_y^1\right),\ \psi\left(s_x^2, s_y^2, \theta^2, c_x^2, c_y^2\right),\ldots,\psi\left(s_x^N, s_y^N, \theta^N, c_x^N, c_y^N\right)\right)$$
$$W = \left(w_1, w_2, \ldots, w_N\right) \tag{5}$$

ein *Gabor-Wavelet-Netzwerk* (Ψ,W) für das Bild b. Die Approximation b´ des Bildes b mittels des erzeugten Gabor-Wavelet-Netzwerks erfolgt durch

$$b' = \sum_{i=1}^{N} w_i \psi_{ni} = \Psi W^T \tag{6}$$

Die Qualität der Approximation b´ hängt von der Anzahl der benutzten Gabor-Wavelets ab, die experimentellen Messungen wurden mit 9, 36, 64 und 81 Gabor-Wavelets durchgeführt. Durch die Optimierung mittels Energieminimierung sollen die Gabor-Wavelets derart positioniert werden, dass eine möglichst gute Approximation von signifikanten Bildbereichen erzielt wird. Beim erwähnten Verfahren zur Gesichtserkennung wird für diesen Zweck der Gradientenabstiegsalgorithmus von LEVENBERG-MARQUAD [16] benutzt. Dieser Algorithmus setzt Grauwertbilder voraus und hat den Nachteil, dass die Optimierung in einigen Fällen die lokalen Minima nicht verlassen kann. Zur Vermeidung dieser Probleme müssen die Startparameter sorgfältig gewählt und an die aktuelle Anwendung angepasst werden. Bei einer Gesichtserkennung ist diese Voraussetzung aufgrund des vorhandenen Kontextwissens erfüllbar. Für allgemeine Anfrageregionen ist der notwendige Aufwand groß und die Eignung der gewählten Parameter nicht immer vorhersehbar. Aus diesem Grund wird für das dynamische Bildretrieval eine modifizierte Variante des Algorithmus von LEVENBERG-MARQUAD mit einer adaptiven Bestimmung der Schrittweite eingesetzt. Anstatt der festen Schrittweite in Richtung des Gradienten werden die neu zu untersuchenden Bereiche in Abhängigkeit vom Optimierungsergebnis bestimmt. Wird eine Verbesserung gegenüber dem aktuellen Optimierungswert erzielt, so wird anschließend die Schrittweite verkleinert, um die lokale Umgebung des Ausschnitts eingehender zu analysieren. Andernfalls wird die Schrittweite vergrößert, um eine alternative Optimierung auszutesten und die lokalen Minima zu überwinden.

Die Vorgehensweise bei der Initialisierung und der Optimierung der Gabor-Wavelets ist an das bekannte Pyramidenschema von LAPLACE angelehnt. Zunächst werden 16 Gabor-Wavelets in einer äquidistanten 4×4-Anordnung zwecks grober Optimierung über die Anfrageregion q verteilt und somit die unterste Pyramidenschicht definiert. Die Optimierung ergibt die approximierte Anfrageregion $q_{4\times4}$. Diese wird anschließend von der Originalvorlage subtrahiert und das Differenzbild $q'=q-q_{4\times4}$ einer weiteren Optimierung unterzogen. Dazu wird die Auflösung verfeinert und ein Netzwerk mit 36 Gabor-Wavelets verwendet, die wiederum in einem äquidistanten 6×6-Gitter angeordnet sind. Die Energieminimierung führt zur Approximation $q_{6\times6}$ des Differenzbildes q'. Die Addition der Approximationen $q_{4\times4}$ und $q_{6\times6}$ ergibt somit die Approximation q_{GWN} der Anfrageregion q. Diese Vorgehensweise kann weiter fortgesetzt werden, bis der Approximationsfehler eine benutzerdefinierte Schwelle unterschreitet. Eine detaillierte Beschreibung dieses Optimierungsschemas gibt beispielsweise [17].

Mit der Bestimmung des GWNs q_{GWN} entsteht eine mathematische Beschreibung für die Anfrageregion q, die für das dynamische Bildretrieval eingesetzt wird. Der aktuelle Bildausschnitt wird als Eingabeparameter verwendet und ein Rekonstruktionsergebnis gemäß des optimierten GWNs erzeugt. Die Summe der Fehlerquadrate gibt die Übereinstimmung zwischen der Anfrageregion und dem aktuellen Bildausschnitt an und dient somit als Ähnlichkeitsmaß. Ein wichtiger

Vorteil der mathematischen Regionbeschreibung besteht darin, dass während der Ähnlichkeitsbestimmung eine zusätzliche Optimierung zwecks Wiederherstellung von Korrespondenzen eingesetzt werden kann. Während die bislang vorgestellten Retrievalmethoden ausschließlich Pixel an den identischen Positionen in den beiden zu untersuchenden Bereichen verknüpften, können bei dieser GWN-basierten Methode dynamische Anpassungen an Variationen bezüglich der Orientierung, Skalierung und Positionierung während der Laufzeit vorgenommen werden. Die Robustheit dieser Invarianz hängt ab von der verwendeten Anzahl von Gabor-Wavelets und deren Skalierung sowie von der Struktur der Anfrageregion. Experimentelle Messungen im Fall der Gesichtsverfolgung zeigten, dass eine Translation von ±10 Pixel in x- und in y-Richtung, Skalierungsunterschiede von 20% und Rotationsveränderungen bis maximal 10° kompensiert werden können [11].

Dieses Verhalten wird durch eine Reparametrisierung des Gabor-Wavelet-Netzwerks während der Laufzeit erzielt. Darunter wird die Anpassung der Parameter (c_x,c_y,Θ,s_x,s_y) an den aktuellen Bildausschnitt verstanden, so dass geringfügige Veränderungen in der Position, Ausdehnung und Rotation ausgeglichen werden können. Dazu wird ein so genanntes Superwavelet [18] benutzt. Ein *Superwavelet* Ψ_n wird definiert als eine lineare Kombination der Gabor-Wavelets ψ_{ni}, d.h.

$$\Psi^{S}_{(s_x,s_y,\theta,c_x,c_y)} = \sum_i w_i \psi_{(s_x^i,s_y^i,\theta^i,c_x^i,c_y^i)} \left(\begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x - c_x \\ y - c_y \end{pmatrix} \right) \quad (7)$$

wobei (s_x,s_y) die Ausdehnung, Θ die Rotation und (c_x,c_y) die Verschiebung festlegen.

SZU ET AL. zeigen in [18], dass das Superwavelet auch eine kontinuierliche, von den Parametern (c_x,c_y,Θ,s_x,s_y) abhängige Waveletfunktion darstellt. Aus diesem Grund kann das Optimierungskriterium wieder eingesetzt werden. Als Rekonstruktionsziel wird diesmal der aktuelle Bildausschnitt a_k verwendet. Die Fehlerenergie *E* mit

$$E = \min_{\forall(s_x,s_y,\theta,c_x,c_y)} \left\| a_k - \Psi^{S}_{(s_x,s_y,\theta,c_x,c_y)} \right\|_2^2 \quad (8)$$

wird so lange minimiert, bis ein vorgegebener Schwellwert unterschritten ist oder eine bestimmte Anzahl von Iterationen durchlaufen wurde. Am Ende dieser Phase werden bei erfolgreicher Optimierung die notwendigen Korrespondenzen zwischen den übereinstimmenden Bereichen in der Anfrageregion q und im aktuellen Bildausschnitt a_k hergestellt. Die anschließende Berechnung der Summe der Fehlerquadrate dient als Ähnlichkeitsmaß. Der Bereich mit der geringsten Summe wird zum ähnlichsten Bereich deklariert und dient als Grundlage für die Einordnung des entsprechenden Bildes in das Ranking.

4.3 Qualitätsmessungen beim Bildretrieval mit Gabor-Wavelet-Netzwerken

Zur Definition des GWNs wurde ein 3×3-Gitter mit Gabor-Wavelets verwendet. In der zweiten Phase wurde das Differenzbild mit einem feineren 3×9-Feld optimiert, so dass die Kombination der beiden zugehörigen Gabor-Wavelet-Netzwerke das

Superwavelet ergab. Während der Laufzeit wurde das Superwavelet an den aktuellen Ausschnitt durch einen Optimierungsprozess angepasst. Dabei wurde der Skalierungsbereich auf 70% bis 140% der Vorlagengröße beschränkt. Die Anzahl durchgeführter Optimierungsiterationen spielte eine untergeordnete Rolle, da ab 15 Durchläufe nur noch geringfügige Abweichungen der Retrievalgüte festgestellt wurden.

Die Suche nach einem Verkehrszeichen oder Logo lieferte stets das gesuchte Symbol an erster Stelle und nachfolgend Symbole mit übereinstimmender Form. Bei den modifizierten Landschaftsbildern wurden hingegen 12 der 14 möglichen Bilder korrekt identifiziert, dies entspricht einer Recallquote von 85.7%. Die Recallquote bei der Anwendung auf die realen Aufnahmen sank auf 56.5%, d.h. es wurden 13 der 23 relevanten Bilder korrekt identifiziert und unter den ersten 32 Treffern zurückgegeben. Die ersten 16 Treffer sind in Abbildung 5 zusammengefasst.

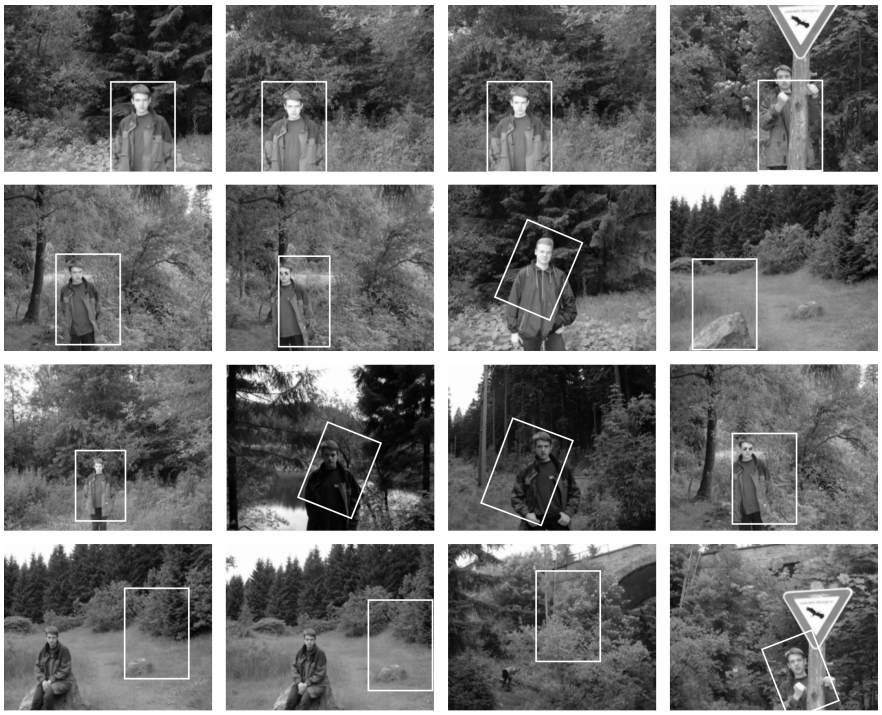

Abbildung 5: Erste 16 Treffer der GWN-basierten Objektsuche in realen Aufnahmen

Eine Analyse dieser Treffer ergibt, dass einige als komplex zu wertenden Einstellungen – beispielsweise die teilweise verdeckte oder verkleinerte Person bzw. die Aufnahme unter schlechten Beleuchtungsverhältnissen – fehlerfrei erkannt wurden. Andererseits beinhalten die meisten der nicht erkannten Bilder eine klar ausgeprägte Abbildung der Person, die zwar eine variierende Körperhaltung einnimmt, aber intuitiv sofort identifizierbar ist. Einige dieser Beispiele sind in Abbildung 6 dargestellt. So unterscheidet sich die Darstellung der stehenden Person

in Abbildung 6(a) bzw. der sitzenden Person in Abbildung 6(b) und Abbildung 6(c) intuitiv kaum von der Originalvorlage. Die größte Übereinstimmung wurde jedoch im texturreichen Hintergrund entdeckt, beispielsweise bei der Steinmaserung oder bei dem von Baumästen verdeckten Kasten.

Bei der Analyse dieses ungewöhnlichen Verhaltens muss jedoch beachtet werden, dass die Farbinformation gänzlich vernachlässigt und die Vorlage durch eine geringe Anzahl von charakteristischen Kernbereichen beschrieben wurde. Viele dieser Kernbereiche befinden sich in Gebieten mit starken Übergängen und werden durch die abweichende Körperhaltung nachhaltig verändert. So sind bei der Vorlage die Hände der Person nicht sichtbar und die Arme ruhen neben dem Körper. In Abbildung 6(b) und Abbildung 6(c) sind die Arme hingegen nach vorne verschränkt und die Jacke dadurch zur Mitte gedrückt und geschlossen. Viele Übergänge zwischen der Jacke und dem Pullover in der Originalvorlage sind somit verschwunden.

(a) (b) (c)

Abbildung 6: Beispiele für nicht erkannte reale Aufnahmen

Diese Abweichungen können bei eher homogenen Hintergründen durch die Übereinstimmung der restlichen Kernbereiche kompensiert werden. Bei den hier benutzten texturreichen Szenarien ist allerdings die Wahrscheinlichkeit groß, ähnliche Übergänge bei den Grashalmen und der Steinmaserung zu finden. Im Lauf des Optimierungsprozesses wird die geringe Anzahl betrachteter Gabor-Wavelets noch zusätzlich an diese Übergänge ausgerichtet, so dass eine gute Übereinstimmung und somit eine geringe Fehlerdifferenz erzielt wird. Diese Probleme können allerdings durch eine Kombination mit anderen, etwa farb- und konturbasierten Verfahren, eliminiert werden. Dadurch wird die Annahme bestätigt, dass das Problem des Bildretrievals nur durch eine Kombination vieler Merkmale – analog zur Verarbeitung der visuellen Eindrücke im menschlichen Gehirn – gelöst werden kann. Dies erfordert jedoch für jede aktuelle Anwendung eine Reihe von langwierigen Experimenten zur Bestimmung der geeigneten Merkmale, Parameter und Gewichte.

5 Zusammenfassung

In diesem Beitrag wurden zwei Methoden für das dynamische Retrieval von Bildern, vorgestellt, welche eine gezielte Suche nach benutzermarkierten Elementen in einem allgemeinen Bildbestand ermöglichen. Durch die manuelle Selektion der

Anfrageelemente können die Beschreibungsressourcen effizienter und präziser eingesetzt werden. Außerdem wird ein Kontextwissen gewonnen, das anschließend in den Retrievalprozess zur Interpretation der untersuchten Bereiche eingebunden wird. Dazu wurden wavelet- und gaborbasierte Methoden entworfen, die mehrere Verfahren zur Umsetzung des dynamischen Bildretrievals enthalten. Insbesondere wurden Abweichungen hinsichtlich affiner Verzerrungen sowie fehlende Korrespondenzen durch einen Optimierungsprozess beseitigt. Eine experimentelle Evaluation mit Bildserien unterschiedlicher Komplexität zeigte, dass bereits mehr als 93% der gewünschten Objekte identifiziert werden können.

Die entwickelten Methoden sind nur ein erster Schritt zur Realisierung eines detailbasierten multimedialen Retrievals. Die Retrievalqualität lässt sich durch die Einbindung von zusätzlichem Kontextwissen und die Kombination mehrerer Merkmale weiter steigern, so dass entsprechende Ansätze für verschiedene Medienklassen untersucht werden müssen.

6 Danksagung

Die Implementierung der vorgestellten Verfahren sowie die Messungen der Retrievalgüte wurden von Herrn Frank Necas-Nießner und Herrn Steffen Priesterjahn durchgeführt.

Literaturverzeichnis

[1] S. Ornager. Image retrieval: Theoretical and empirical user studies on accessing information in images. *Proceedings of the 60th Meeting of the American Society for Information Science, volume 34, pages 202–211, 1997.*

[2] O. Kao, I. la Tendresse. On the Impact of the Number of Coefficients on the Quality of Wavelet-based Image Retrieval. *Proceedings of the International Conference on Computer Imaging Science, Systems and Technologies, pages 361–372, 2001.*

[3] S. Mallat. *A theory for multiresolution signal decomposition: the wavelet representation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 11:674–693, 1989.

[4] J.L. Starck, A. Bijaoui, I. Valtchanov, F. Murtagh. A combined approach for object detection and deconvolution. *Astronomy and Astrophysics Supplement Series, 147:1–10,* 2000.

[5] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, T. Poggio. Pedestrian detection using wavelet templates. *Proceedings of the International Conference on Computer Vision and Pattern Recognition, pages 193–199, 1997.*

[6] C. E. Jacobs, A. Finkelstein, D. H. Salesin. Fast multiresolution image querying. *Proceedings of ACM SIGGRAPH, pages 277–286, 1995.*

[7] O. Kao, I. La Tendresse. CLIMS - A system for image retrieval by using colour and wavelet features. *Advances in Information Systems, LNCS 1909, pages 238 – 248, 2000.*

[8] M. Feil, A. Uhl. *Real-time image analysis using wavelets: the à trous algorithm on MIMD architectures*. IS&T/SPIE's Electronic Imaging Newsletter, 9(2):4–5, 1999.

[9] F. Necas-Niessner. *Allgemeine Objektsuche für Bilddatenbanken mittels Wavelet-Analyse zur dynamischen Objekterkennung*. Diplomarbeit, TU Clausthal, 2001.

[10] M.J. Lyons, A.Plante, S. Jehan, S. Inoue, S. Akamatsu. Avatar creation using automatic face processing. *Proceedings of ACM Multimedia 98, pages 427–434, 1998.*

[11] V. Krüger, G. Sommer. *Gabor wavelet networks for object representation*. Technical Report 2002, University of Kiel, 2000.

[12] D. H. Hubel, T. N. Wiesel. *Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex*. Journal of Physiology, 160:106–154, 1962.

[13] T.S. Lee. *Image representation using 2D gabor wavelets*. IEEE Transaction on Pattern Analysis and Machine Intelligence, 18(10):959–971, 1996.

[14] R.P. Würtz. *Multilayer Dynamic Link Networks for Establishing Image Point Correspondences and Visual Object Recognition*. Dissertation, Ruhr-Universität Bochum, Verlag Harri Deutsch, 1995.

[15] J.C. Vorbrüggen. *Zwei Modelle zur datengetriebenen Segmentierung visueller Daten*. Dissertation, Ruhr-Universität Bochum, Verlag Harri Deutsch, 1995.

[16] J. More. The Levenberg-Marquardt algorithm: Implementation and Theory. *Numerical Analysis, LNM 630, pages 105–116, 1977.*

[17] V. Krüger. *Gabor Wavelet Networks for Object Representation*. Dissertation, Christian-Albrecht University, Kiel, 2001.

[18] H. Szu, B. Telfer, S. Kadambe. *Neural network adaptive wavelets for signal representation and classification*. Optical Engineering, 31(9):1907–1961, 1992.